Intelligent Systems Technologies to Assist in Utilization of Earth Observation Data

Gail McConaughy, Code 586

IA-KBS Team:

- H. (Rama) Ramapriyan*, G. McConaughy*, S. Morse**, D. Isaac***, LiPing Di****, Wen Li****, Chris Lynnes*

- * NASA Goddard Space Flight Center, Greenbelt, MD

- ** SoSA Corporation, Chantilly, VA

- *** Business Performance Systems, Bethesda, MD

- **** George Mason University

- ISD Technology Exchange Workshop

- January 25, 2005

- Large and growing data collections from the Earth Observing System
    - 3.4 petabytes of data
    - 48 million files
    - 3.5 terabytes/day accumulation
- Distributed, heterogeneous data systems
    - ~70 data centers
    - Complex "value chains"
- Broad & diverse user community
    - Research, applications, education
- Limited human capacity to examine large volumes of data
   - Users need information, not just data

Widely Distributed and Heterogeneous Data

- Over 70 NASA funded Earth science "data centers" across the US, plus interagency and international partners

- Trend is for further distribution

  - Recent cooperative agreements add to the network of PI systems (REASoN)

◆ Distributed Active Archive Centers (DAACs)
● Science Investigator-led Processing Systems
▲ REASoN Projects
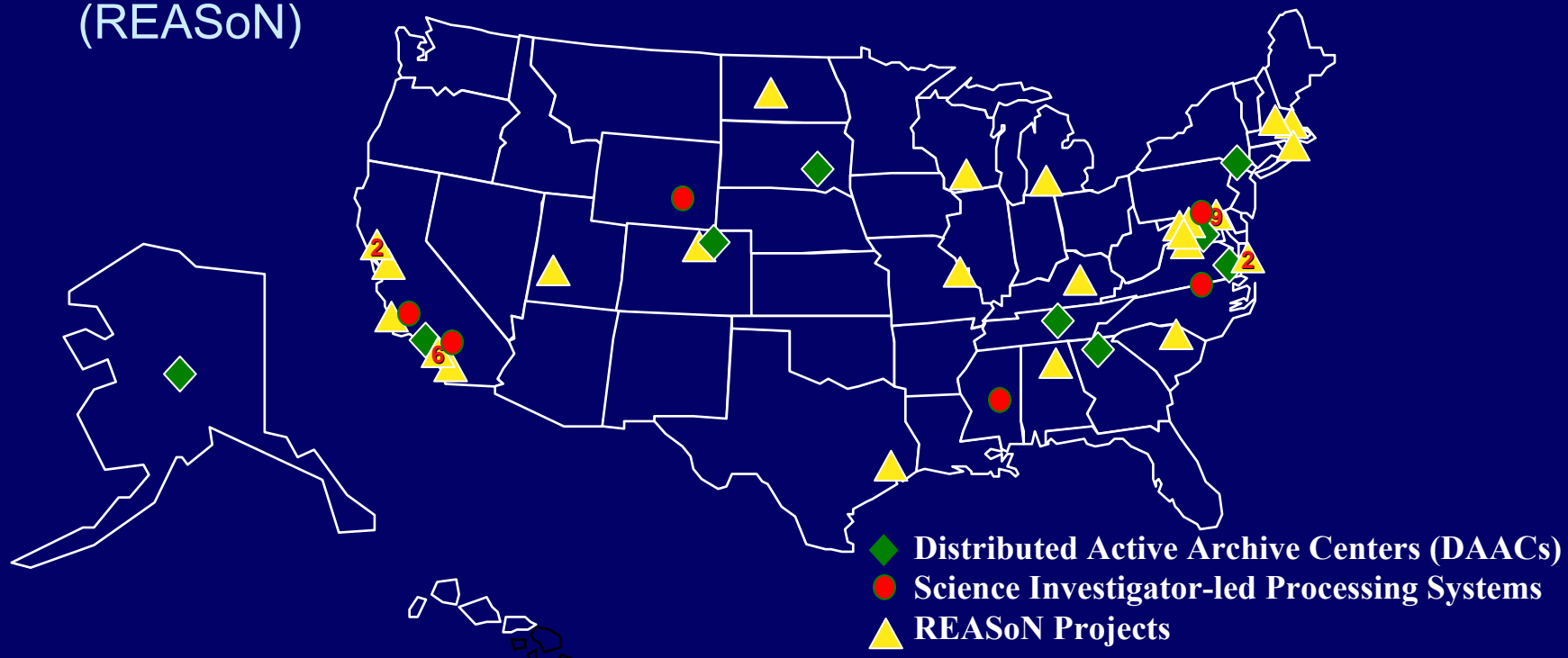
Developed by: IAS-KBS Team: Ramapriyan, McConaughy, Morse, Isaac, Di, Li, Lynnes

The Challenge: Data Utilization Issues

- Timeliness
  - New applications require near-real-time data delivery
  - Human-based data quality assessment can take weeks or longer

- Access
  - Users need more assistance in locating relevant data in large archives
  - Content-based metadata and indexes could help

- Understandability
  - Users need a concise description of the salient characteristics of data
  - But, current data systems are generally oblivious to the content

- Readiness for Use
  - Users want information, not just data
  - Need to move up the data → information → knowledge chain

- Responsiveness
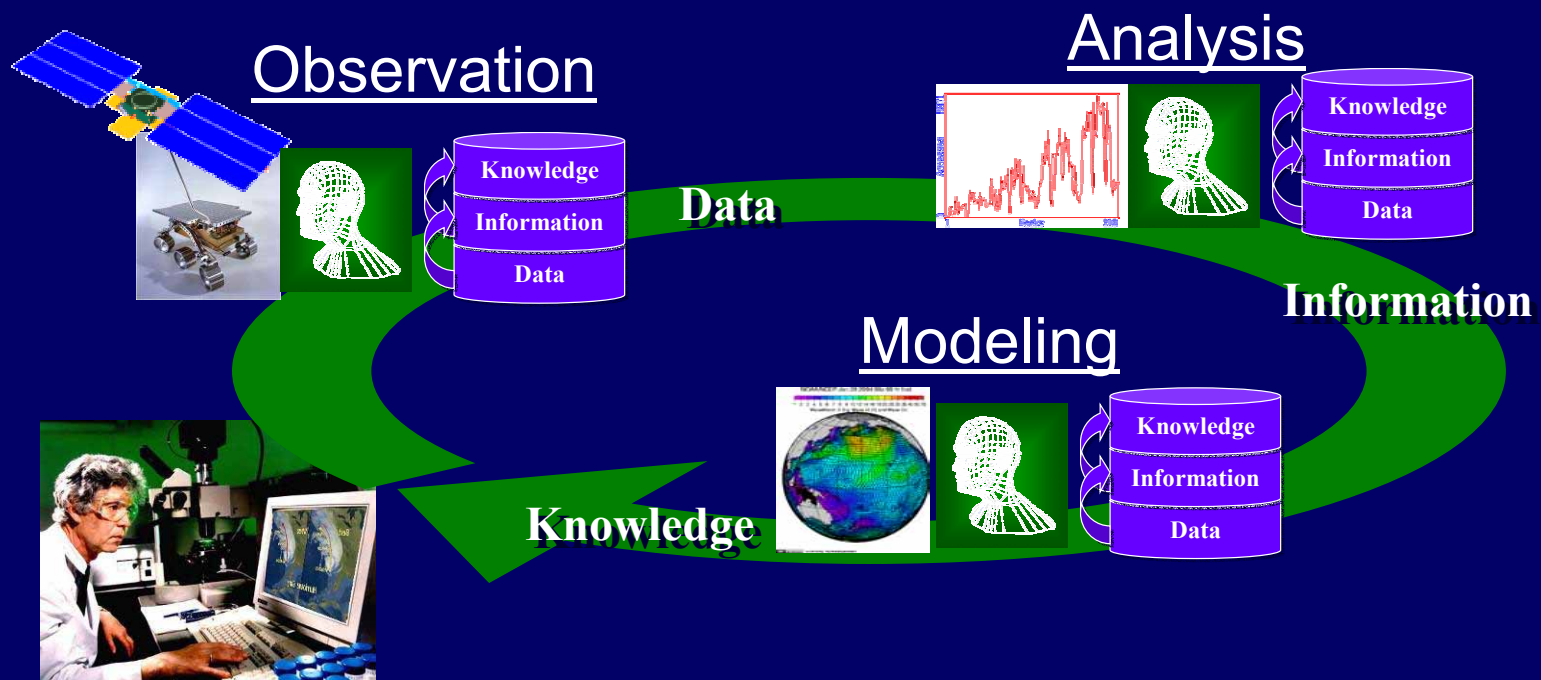 - Systems should be aware of user needs and adapt to them

Developed by: IAS-KBS Team: Ramapriyan, McConaughy, Morse, Isaac, Di, Li, Lynnes

Intelligent Archives in the Context of Knowledge Building Systems (IA-KBS)

- Data archives exist throughout the information value chain

- Intelligence with feedback loops makes systems more effective

- Distributed intelligent components collaborate to achieve user goals

Developed by: IAS-KBS Team: Ramapriyan, McConaughy, Morse, Isaac, Di, Li, Lynnes
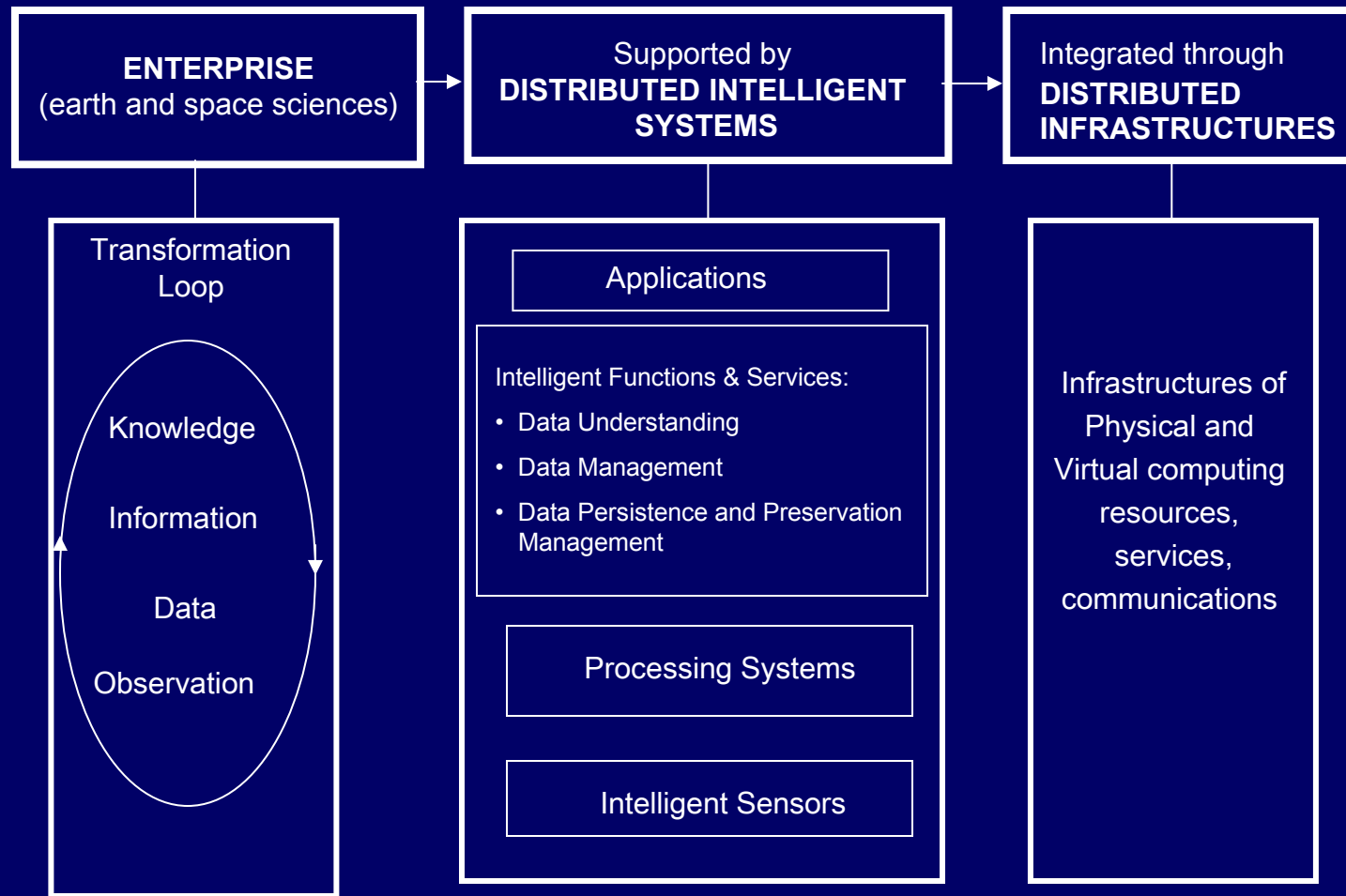
The Opportunity

- Data mining algorithms
    - Induction of general characteristics, relationships, & patterns from specific data
    - Successfully moved from labs to industrial use

- Intelligent data understanding
    - Research sponsored by NASA's Intelligent Systems Project
    - 22 research projects exploring a variety of algorithms applied to a variety of data...<u>including remote sensing data</u>

- Affordable high-performance computing
    - Improvements may make large-scale data mining feasible
    - Grid technologies could also provide needed capacity

Developed by: IAS-KBS Team: Ramapriyan, McConaughy, Morse, Isaac, Di, Li, Lynnes

ENTERPRISE
(earth and space sciences)

Supported by
DISTRIBUTED INTELLIGENT SYSTEMS

Integrated through
DISTRIBUTED INFRASTRUCTURES

Transformation Loop

Knowledge

Information

Data

Observation

Applications

Intelligent Functions & Services:
- Data Understanding
- Data Management
- Data Persistence and Preservation Management

Processing Systems

Intelligent Sensors

Infrastructures of Physical and Virtual computing resources, services, communications

Developed by: IAS-KBS Team: Ramapriyan, McConaughy, Morse, Isaac, Di, Li, Lynnes

IA-KBS – Relevant Technologies

- Distributed system architectures

    - Especially, Grid technologies

- Intelligent data understanding algorithms

    - Fern & Brodley: understanding high-dimensionality data using clustering, re-projection, cluster ensembles

    - Kumar et al: discovering climate indices using clustering on time-series data

    - Teng: identifying and removing anomalies to improve classifier performance

    - Kargupta: extending data mining algorithms to distributed architectures

    - Smelyanskiy: Bayesian inference of non-linear dynamical model parameters

    - Nemani & Golden: dynamic assembly of data and operators to satisfy a user's information goal

    - LeMoigne: sub-pixel accurate image registration for data fusion

- Promising data mining algorithms have been identified and applied to remote sensing data in a laboratory environment

- Design and implementation of a large-scale data mining experiment
    - Select and port IDU algorithm to large-scale storage system connected via grid technology to distributed processing capabilities

- Next step is to demonstrate utility and scalability in an operational environment

- Address test bed that would be generally useful to any Decision Support System Environment